A Multifaceted Approach to Understanding the Botnet phenomenon

Javier Zugasti Raposo

30th July 2007

Technische Universität Berlin

Seminar “Internet Measurements”

Betreuer: Gregor Maier

Contents

- Introduction
- Botnets: Global view
- Measurements
  - Malware Collection
  - Binary Analysis
  - Tracking of Botnets
- Results and Analysis
  - Prevalence of Botnet Phenomenon
  - Spreading and growth patterns
  - Effective Botnet sizes
  - Taxonomy
- Conclusion
- References

Introduction

- aim to clarify mysteries within botnets.

- Botnets:

<u>What?</u> Networks of infected hosts (bots) controlled by a person (botmaster).

<u>How?</u> IRC protocol

<u>What for?</u> Extortion of Internet businesses, identity theft, spamming, software piracy,...

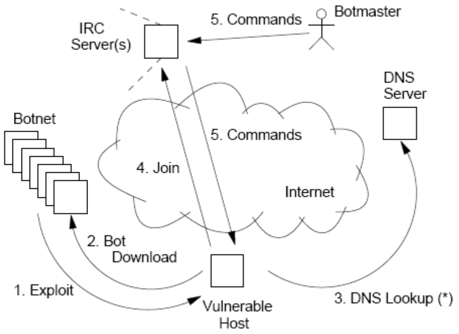
- Development of a multifaceted and distributed measurement infrastructure.

Steps: 1.Malware collection
2.Binary Analysis
3.IRC- and DNS-tracking

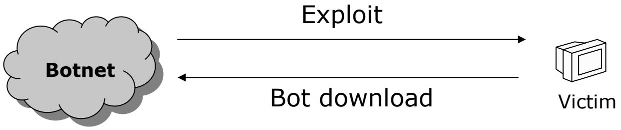
- Analysis of results: High representation in the overall malicious attemps (27%).
Great evidence in DNS domains(11%).

Botnets: Global View

Armies of bots commanded by a botmaster.
How does the infection process occur?
How do the bots contact the botmaster?



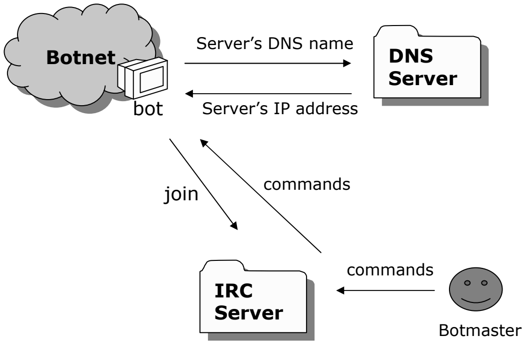
1. Infection process



-Infection strategies common to other kinds of malware (self replicating worms, email, …).
-Convince the victim to execute code.

-Code is executed and the bot binary downloaded.
-Binary installed in the background and started after every reboot.

2. DNS lookup and joining



-IRC server's name resolved.

-join:3 steps of authentication: bot->server
bot->master
botmaster->bot

-bot executes default commands.

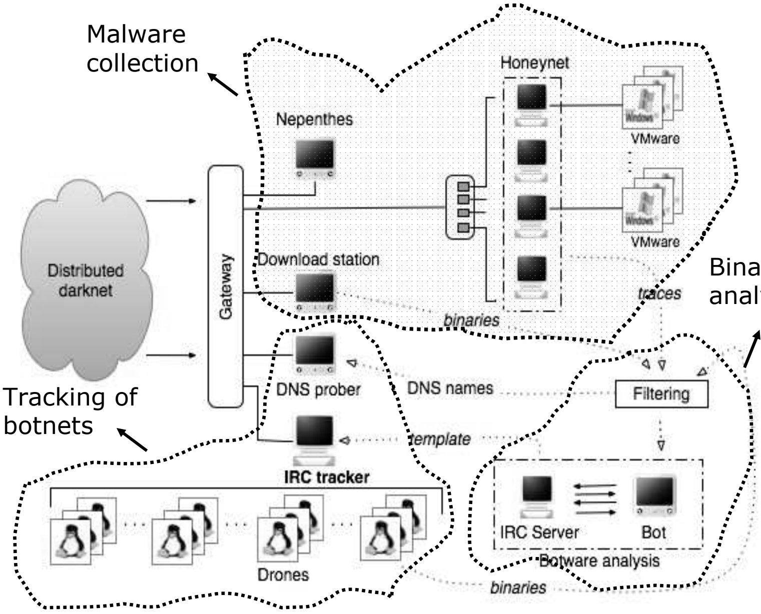
Measurements

Three phases in the measurement procedure:

1. Malware collection: *Nephenthess platform*, honeynet and download station.

2. Binary analysis

3. Tracking of Botnets.
   - IRC tracking
   - DNS tracking



Malware collection

Binary analysis

Tracking of botnets

1. Malware collection

A distributed *darknet* is used and results are extrapolated to the Internet.

14 nodes with access to the darknet.

Modified version of *Nephenthess platform*:

- Mimics the answers generated by the victims to collect shellcodes.
- List of URLs (contained in the binaries) to be downloaded.

Honeynet to compliment Nephenthess:

- Honeypots run unpatched WinXP versions.
- Establish IRC connections.
- Compared to clean XP images.

Gateway to engage all parts (NAT).

2. Binary Analysis

Analysis tool used for the analysis and extraction of binaries features.

-Network level analysis:

Bots are run in a controlled environment and traffic logs are processed on a created server.

Network-Fingerprint: Targets of DNS requests, destination IP-addresses, ports. Also whether or not scanning occured.

-Application level analysis:

IRC server is runned and set to listen on the ports. Bots connect to it.
Fingerprint: password, nickname, mode and channels to be joined

With both fingerprints and botnet’s dialect it is possible to join a botnet in the wild.

3. Tracking of botnets

Performed in two different ways:

-IRC tracker:

Development of IRC client to join the IRC channel.

Pretends to follow all commands from the botmaster.

To appear real pre-filtering is required (suppresion of information).

-DNS tracker:

As bots ussually send DNS requests, a large number of DNS servers are probed to find Evidence in their cache.

Cache hits as measurements.

Results and Analysis

Results include traffic captured at the darknet, IRC logs and DNS cache hits.

The most interesting results are on:

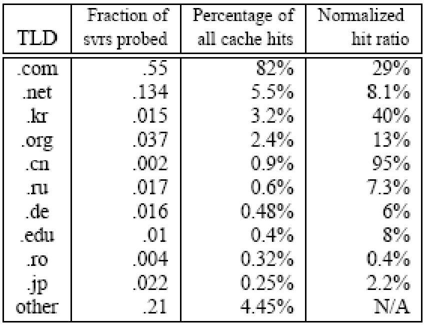
- Prevalence of the botnet phenomenon.
- Spreading and growth patterns.
- Effective botnet sizes.
- Taxonomy.

1. Prevalence of the botnet phenomenon

Prevalence results extracted from DNS probing.

11% of the total amount of servers showed at least one cache hit.

Statistics of DNS servers supporting clients involved in at least one botnet:



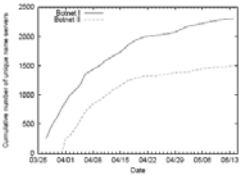
TLD	Fraction of svrs probed	Percentage of all cache hits	Normalized hit ratio
.com	.55	82%	29%
.net	.134	5.5%	8.1%
.kr	.015	3.2%	40%
.org	.037	2.4%	13%
.cn	.002	0.9%	95%
.ru	.017	0.6%	7.3%
.de	.016	0.48%	6%
.edu	.01	0.4%	8%
.ro	.004	0.32%	0.4%
.jp	.022	0.25%	2.2%
other	.21	4.45%	N/A

Example:

-55% of the probed servers were *.com*
-It registered 82% of all cache hits detected.
-29% of *.com* servers probed Registered at least a hit.

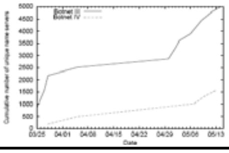
2. Spreading and growth patterns

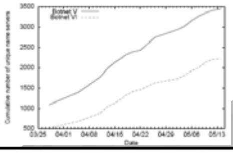
Spreading methods:mail, web and active scanning.
Scanning is the most effective.



2 types of scanning:
-Worm-like botnets:
Continuous scan on some ports.
Semi-exponential growth pattern.

-Botnets that vary their ways of scanning:
Localized,uniform…
Difficult to track because of their intermittent behaviour.
Staircase or linear growth patterns.



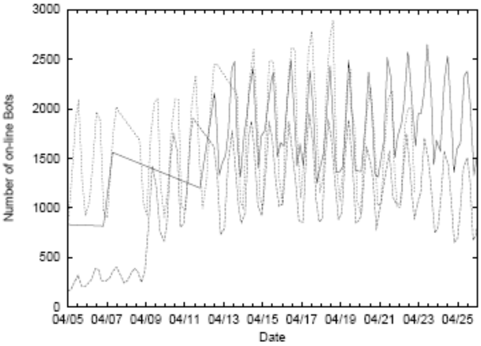


3. Effective botnet sizes

Effective size: amount of botnets connected to a channel at the same time.

Maximum size of the online population smaller than fingerprint’s size.

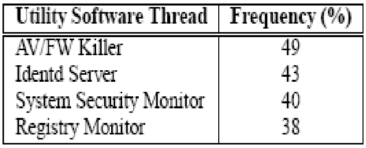
On-line bots for 3 different botnets



Average footprints greater than 10,000 while at most 3000 bots online at the same time.

4. Taxonomy

What processes do bots run?



Utility Software Thread	Frequency (%)
AV/FW Killer	49
Identd Server	43
System Security Monitor	40
Registry Monitor	38

Are anti-viruses scanners prepared?

ClamAV: 137/192
Norton: 179/192

Conclusion

- Severe threat to the Internet.
- Little knowledge of their behaviour.
- IRC because of its versatility.
- Variable effective sizes.
- More and more sophisticated.

Thank you for your attention…

References

[1] Original paper: A Multifaceted Approach to Understanding the Botnet Phenomenon. Moheeb Abu Rajab, Jay Zarfoss, Fabian Monrose, Andreas Terzis. John Hopkins University

[2] The Nephenthes Platform. www.mwcollect.org

[3] The UnrealIRC Team. www.unrealircd.com

[4] Honeynet Project and Research Alliance. www.honeynet.org/papers/bots